

Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning†

Year: 2020 | Citations: 101 | Authors: Amol Thakkar, Veronika Chadimová, E. Bjerrum, O. Engkvist, J. Reymond

Abstract

Computer aided synthesis planning (CASP) is part of a suite of artificial intelligence (AI) based tools that are able to propose synthesis routes to a wide range of compounds. However, at present they are too slow to be used to screen the synthetic feasibility of millions of generated or enumerated compounds before identification of potential bioactivity by virtual screening (VS) workflows. Herein we report a machine learning (ML) based method capable of classifying whether a synthetic route can be identified for a particular compound or not by the CASP tool AiZynthFinder. The resulting ML models return a retrosynthetic accessibility score (RAscore) of any molecule of interest, and computes at least 4500 times faster than retrosynthetic analysis performed by the underlying CASP tool. The RAscore should be useful for pre-screening millions of virtual molecules from enumerated databases or generative models for synthetic accessibility and produce higher quality databases for virtual screening of biological activity.